

Gamma/Proton separation study for the LHAASO-WCDA detector

Wenyang Liao^[1], Min Zha^[2], Wande Fan^[1], Chunxu Yu^[1] and Minjun Chen^[2] on
behalf of LHAASO collaboration

[1]University of Nankai, Tianjin 300071, China

[2]Institute of High Energy Physics, Chinese Academy of Sciences, Beijing, 100049, China

E-mail: zham@ihep.ac.cn

The Large High Altitude Air Shower Observatory (LHAASO) is going to be built at an altitude of 4410 meters in Daocheng, Sichuan Province, China. The Water Cherenkov Detector Array (WCDA), one of the major components of the LHAASO experiment, focuses on surveying the Northern sky for gamma ray sources in a wide energy range (0.1 to 30 TeV). One of the main tasks of the data analysis of the WCDA is to suppress the large number of background events originated from primary cosmic rays. Rather than a single key parameter for Gamma/Proton discrimination, 4 sensitive parameters are chosen for the purpose of further improving the detector sensitivity, using some multi-variate analysis methods such as the artificial neural network and the boosting decision tree. By analyzing the simulation data, these two multi-variate analysis algorithms both manifest excellent Gamma/Proton separation powers, and greatly improve the quality factor and the sensitivity of the WCDA at the low energy.

35th International Cosmic Ray Conference – ICRC2017
12-20 July, 2017
Bexco, Busan, Korea

1. Introduction

Gamma rays, especially very high-energy(VHE) cosmic gamma rays insides or beyond our galaxy, attracted more and more scientists attention because of its electric neutrality providing more information about their original source. LHAASO-WCDA is a ground-based air-shower array with the 78,000 m² coverage area, to be constructed at high altitude of 4410 meters a.s.l in Daocheng, Sichuan Province, China [1]. It will focus on surveying the northern sky for steady and transient sources from 0.1 TeV to 30 TeV [2], with the advantages of a desirable angular resolution, wide field of view and low threshold energy.

Comparing with gamma ray sources, cosmic ray events produce high-level background. In order to achieve good performance on primary gamma ray detection, it is critical for LHAASO-WCDA to find an effective method to reject the background of cosmic rays. The target of this paper is to achieve good gamma/proton separation power. The paper is organized as follows: the LHAASO-WCDA array is introduced in Section 2, and details about MonteCarlo (MC) simulation are presented in Section 3. In the section 4, the sensitive parameters on gamma/proton separation are presented. The subsequent Section provide results and analysis of different separation method from a full MC simulation. Finally, we conclude with a discussion and a summary.

2. Water Cherenkov Detector Array

LHAASO-WCDA is planned to be built at Mountain Haizishan with the altitude of 4410 m, Daocheng, Sichuan Province, China. The layout scheme of WCDA detector is shown in Fig. 1. WCDA comprises three water ponds with an effective water depth of 4 m, two of them will cover the area of 150 m×150 m, and the other 300 m×110 m. Each pond is further divided into cells with the size of 5×5 m, partitioned by black plastic curtains to prevent the crosstalk neighboring cells. Every cell has an upward-facing photomultiplier tube (PMT) at the bottom which can collect Cherenkov lights generated by secondary charged particles from extensive air showers. Using the hitting information and coverage, the WCDA becomes very sensitive to VHE gamma-rays and cosmic-rays in the energy range 0.1 TeV-30 TeV, with the advantages of a desirable angular resolution, wide field of view and lower threshold energy.[2]

3. Simulation details

An extensive Monte Carlo simulation had been carried out to simulate the air shower cascade development in the atmosphere using the package of CORSIKA[3] and the response in the LHAASO-WCDA detector based on GEANT4 package. In the simulation, the detector performances, such as time resolution, electronic noises, based on the experimental data, are taken into account.

26 primary elements, with Z=1-26, i.e H-Iron, are simulated to account for a realistic chemical composition of primary cosmic rays. The energy spectrum of individual primary cosmic ray is chosen according to the paper[4]. And the energy spectrum of gamma ray is used from HEGRA measurement[5]. Concerning the detector trigger efficiency, the energy range of the incident cosmic ray events is taken from 50 GeV to 100 TeV. The energy range of gamma ray is starting from

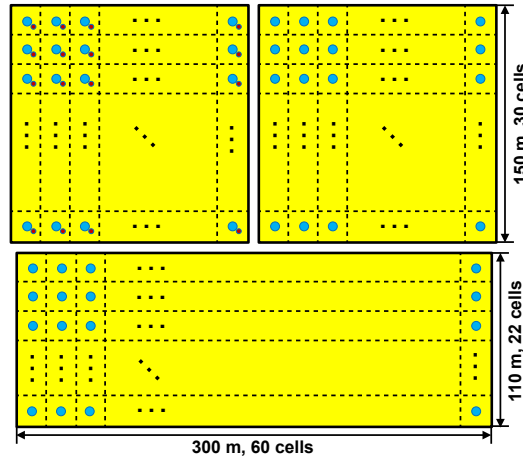


Figure 1: Schematic plot the WCDA.

20 GeV to 100 TeV. The directions of primary cosmic ray events are generated along the Crab Nebular's orbit within a opening angle of 7° around the center, as to gamma signal, the direction is completely generated along the Crab's orbit.

4. The sensitive parameters on gamma/proton separation

Gamma rays induced showers are predominantly electromagnetic, develop primarily via pair-production and bremsstrahlung. The showers are characterized by compact, smooth lateral shower profiles. However, hadronic cosmic rays induced showers, caused by nuclear interactions, are dominated by large fluctuations subsequently decay to pions. Pions with large transverse momenta decay to gamma rays and muons. These sub-showers leads to the hadronic shower has a cluttered, irregular and wide lateral shower profiles.

According to the lateral and longitudinal shower profiles of the secondary particles in the EAS induced by different primaries, we found following four sensitive parameters used as single standard gamma/proton separation methods, and their distribution are shown in Fig.2.

- The first parameter is named as compactness, C , where $C = nFit/cxPE_{45}$ and $nFit$ is the number of fired PMTs during the reconstruction process and $cxPE_{45}$ is the maximum signal outside an exclusion radius of 45 m in each event. From the simulation, the value of $cxPE_{45}$ is expected to be larger for nuclei induced background events due to muon component and the structure of sub-showers in hadronic showers.
- The second parameter is the average density outside of 40m relative to shower core position, named as ρ_{40} .
- As we know, hadronic cosmic rays air showers show sharp lateral distribution comparing with gamma ray showers. Based on this aspect the third parameter is Density Ratio(DR), it is the ratio of average density at two radial distance, 50 m and 10 m.

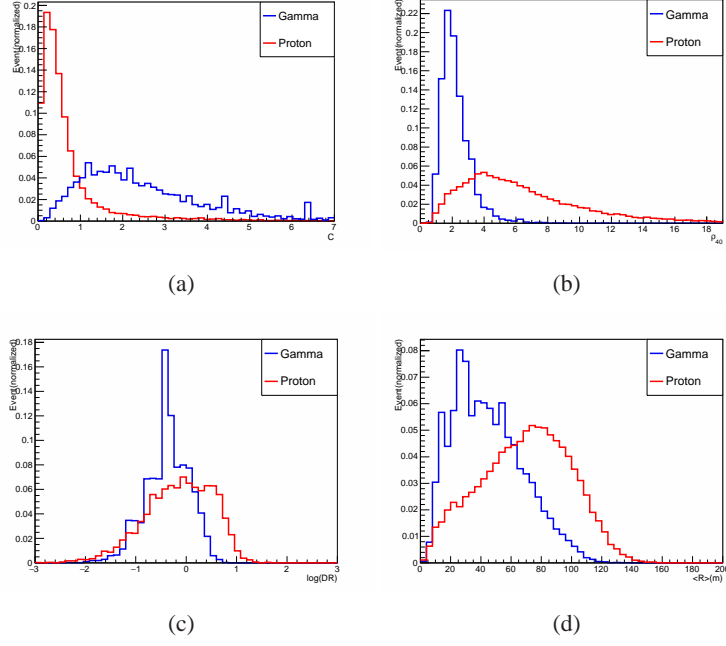


Figure 2: Distribution of sensitivity parameters for gamma and proton initiated events for the multiplicity interval $10 < nFit < 20$.

- The final parameter is the mean lateral spread radius of particle flow from the shower core position ($\langle R \rangle$), a parameter weighted by the number of fired signal, where $\langle R \rangle = \Sigma(N_k r_k) / \Sigma N_k$, where N_k and r_k are the number of detected signal and the radial distance to the core for the k^{th} detector respectively.

5. Separation result

Generally, the performance of the gamma/proton rejection can be evaluated by the quality factor (Q) which is defined as: $Q = \varepsilon_\gamma / \sqrt{\varepsilon_p}$, where $\varepsilon_\gamma(\varepsilon_p)$ is the fraction of gamma (proton) events classified correctly by cut conditions, also called gamma (proton) cut efficiency. Q factor Optimizing will improve the separation of two event classes. Since maximizing quality factor may reduce the gamma ray retention capability, we require $\varepsilon_\gamma > 0.5$ to retain as many signal events as possible after cuts.

The angular resolution and background rejection of WCDA detector are closely correlated with energy, and energy is related to the number of nFit. Therefore, the MC data sample is divided into 7 multiplicity intervals by the number of nFit. A quantitative overview of the separation achieved for different gamma/proton subsamples is summarized in Tab. 1. Optimizing the cut-value of C and ρ_{40} parameters give the very high Q_{max} with background efficiency of 0.2% in the sensitive energy range for WCDA detector. Utilizing the muon component and the sub-showers structure, C and ρ_{40} parameters can be good candidates for single key gamma/proton separation methods.

However, it still exists some shortcomings. The separation power in low energies is poor, which remains 10 percents of background events after cuts. Therefore, we apply using multi-variate analysis method to optimize cut with all sensitive parameters. Tab. 2 presents the correlation matrix

Table 1: Characteristics of the simulated data sample and the value of Q factor for gamma/proton separation of WCDA in the whole range of multiplicity $10 < nFit < 3600$.

nFit	$E_{m,\gamma}/TeV$	$E_{m,p}/TeV$	$Q(DR)$	$Q(<R>)$	Q(C)	$Q(\rho_{40})$	Q(ANN)	Q(BDTG)
10 – 20	0.16	0.53	1.23	1.43	2.19	2.16	4.33	4.61
20 – 50	0.28	0.67	1.31	1.55	2.52	2.51	6.93	7.21
50 – 100	0.62	1.22	1.47	1.62	4.32	4.33	13.78	14.84
100 – 200	1.25	2.35	1.64	1.63	6.43	8.44	18.71	20.45
200 – 400	2.57	4.58	1.69	1.56	8.71	10.45	20.29	23.69
400 – 800	5.99	10.16	1.58	1.49	11.80	10.69	16.94	18.99
> 800	17.64	27.49	1.43	1.44	11.41	8.28	14.57	14.57

between each parameters. There is a sufficient scope for using the differences in the correlation between four sensitive separation parameters for developing multi-variate gamma/proton segregation methodologies.

In this paper, we apply two kinds of machine learning algorithm as multi-variate analysis methods for gamma/proton separation, one that is the Article Neutral Network(ANN) and the other is Boosted Decision Trees with a Gradient boosting algorithm(BDTG). These multi-variate analysis algorithms are implemented and optimized by using the TMVA package inside the ROOT program.

Table 2: Correlation matrix for simulated data sample

		C	ρ_{40}	$<R>$	DR			C	ρ_{40}	$<R>$	DR
gamma	C	1.00	-0.43	-0.27	-0.23	proton	C	1.00	-0.31	-0.17	-0.09
	ρ_{40}	-0.42	1.00	-0.21	0.12		ρ_{40}	-0.31	1.00	-0.03	0.18
	$<R>$	-0.05	-0.21	1.00	0.12		$<R>$	-0.17	-0.03	1.00	0.04
	DR	-0.10	0.12	0.12	1.00		DR	-0.9	0.18	0.04	1.00

ANN is a computational or mathematical model, based on biological neural networks, and a delicate method in multi-variate analysis algorithms [7, 8]. ANN is self-adaptive and data driven methods and can automatically adjust accordingly to the given data. Besides, it is able to estimate the posterior probability which provides the basis of establishing classification rule and performing statistical analysis. A newly developed neural network, named as Multi-layer Perception Artificial Neural Networks, is used in the study. Compared with rectangular cut or likelihood discriminants, BDTG classifier considers nonlinear correlations between input features as well as ANN algorithm. Furthermore, the BDTG increases the performance of classifier and stabilizes the response of the decision trees with respect to fluctuations in the training sample[9]. Output histograms of ANN and BDTG classifiers are shown in Fig.3 for the test sample with the BDTG trained in the multiplicity interval $10 < nFit < 20$, demonstrating the excellent classification power in terms of gamma/proton separation.

6. Comparison of different separation methods

As we know, it is difficult in event reconstruction in the lowest nFit interval where is abundance of data, which results in poor gamma proton separation capability. Using multi-variate analysis method based on ANN and BDTG classifiers can deal with this defect (as is shown in Tab. 1).

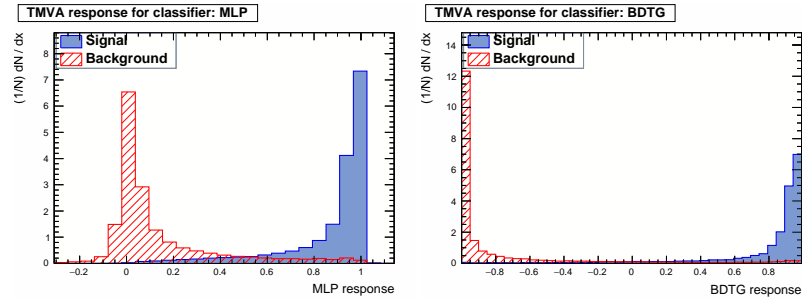


Figure 3: MLPANN(left) and BDTG(right) output for gamma (blue line) and background (red line) events.

For example, in the multiplicity interval $10 < nFit < 20$, by optimizing the cut value of ANN and BDTG output, their maximum Q factor are 4.33 and 4.61, respectively, which can reject more than 98% of background events and remain 60% exceed signal events. Fig.4 shows their receiver operating characteristic(ROC) which shows the true positive rate versus the false positive rate for the different possible cut points of hypothesis test [10]. Furthermore, in the TeV energies, the value of Q factor for these two multi-variate analysis algorithms are more than 3 times bigger. The value of Q factor for BDTG classifier can reach up to 23.69, and it can reject 99.90% of proton background events while remain more than 50% gamma signal events. It demonstrates the excellent classification power of multi-variate analysis methods in terms of gamma/proton separation

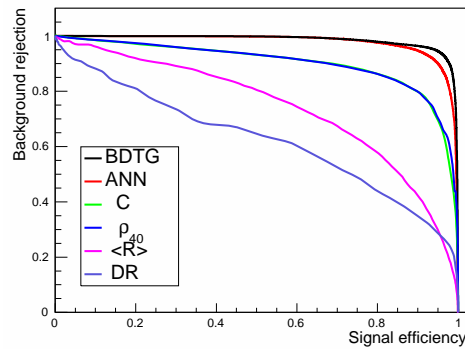


Figure 4: The background rejection versus signal efficiency (ROC curve) obtained by cutting on different methods for the multiplicity interval $10 < nFit < 20$.

7. Conclusion

LHAASO-WCDA detector is a high real-time air-shower array used to study both point-like and extended gamma-ray sources. This study has been demonstrated that the detector has an excellent performance on rejection of hadron ray background. By employing, ANN and BDTG multi-variable analysis algorithms with four sensitivity parameters, WCDA detector can achieve a high value of quality factor in the whole range of multiplicity $10 < nFit < 3600$. Besides, using multi-variable analysis separation methods can greatly improve on single key separation param-

ters in low energies where is abundance of data and difficulties in fitting the shower parameters. It will be able to enhance the detector sensitivity to a given source outside the galaxy.

Acknowledgments

This work is supported in China by NSFC NO.11635011, NO.11675187, the Key Laboratory of Astrophysics and Cosmic Ray, Institute of High Energy Physics, CAS.

References

- [1] Z. Cao et al, Proceedings of 33th ICRC (2013).
- [2] M.J. Chen et al, "Status of Water Cherenkov Detector Array of LHAASO Project", This proceeding.
- [3] D. Heck, J. Knapp, J. Capdevielle, G. Schata, and T. Thouw, "CORSIKA: A Monte Carlo code to simulate extensive air showers", Report FZKA 6019 (1998).
- [4] Jorg R. Horandel. On the knee in the energy spectrum of cosmic rays. *Astroparticle Physics* 19 (2003) 193-220.
- [5] A. Konopelko, F. Aharonian et al. Detection of gamma rays above 1 TeV from the Crab Nebula by the second HEGRA imaging atmospheric Cherenkov telescope at La Palma. *Astroparticle Physics* 4 (1996) 199-215.
- [6] Min Zha, *Nuclear Physics B (Proc. Suppl.)* **175+176**, (2008) 443.
- [7] R.K. Bock, et al. *NIM A* 588 (2008) 424.
- [8] M.D Richard, L. Lippmann, *Neural Computation* 3(1991) 461.
- [9] J.R. Quinlan *Bagging, Boosting, and C4,5* In Proceedings of the Thirteenth National Conference on Artificial Intelligence, (1996).
- [10] T. Fawcett, An introduction to ROC analysis, *Pattern Recognition Letters* 27 (2006) 861-874.